

# A unified recovery bound estimation for noise-aware $\ell_q$ optimization model in compressed sensing

Zhi-Long Dong, Xiao-Qi Yang, Yu-Hong Dai

**Abstract**—In this letter, we present a unified result for the stable recovery bound of  $\ell_q$  ( $0 < q \leq 1$ ) optimization model in compressed sensing, which is a constrained  $\ell_q$  minimization problem aware of the noise in a linear system. Specifically, without using the restricted isometry constant (RIC), we show that the error between any global solution of the noise-aware  $\ell_q$  optimization model and the ideal sparse solution of the noiseless model is upper bounded by a constant times the noise level, given that the sparsity of the ideal solution is smaller than a certain number. An interesting parameter  $\gamma$  is introduced, which indicates the sparsity level of the error vector and plays an important role in our analysis. In addition, we show that when  $\gamma > 2$ , the recovery bound of the  $\ell_q$  ( $0 < q < 1$ ) model is smaller than that of the  $\ell_1$  model, and the sparsity requirement of the ideal solution in the  $\ell_q$  ( $0 < q < 1$ ) model is weaker than that of the  $\ell_1$  model.

## I. INTRODUCTION

The problem of finding the sparsest representation possible in an overcomplete dictionary  $A$  is

$$(P_0) : \begin{aligned} & \min_x \|x\|_0 \\ & \text{s.t. } y_0 = Ax, \end{aligned} \quad (1)$$

where  $y_0 \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$  ( $m \ll n$ ),  $x \in \mathbb{R}^n$ ,  $\|x\|_0$  means the number of nonzero entries in  $x$ . In this letter, we assume that problem (1) always has a unique ideal sparsest solution  $x_0$ , and the sparsity level defined by  $N = \|x_0\|_0$  is small. This problem has applications in many areas, such as statistics and signal processing, etc [1]. However, it is well known that problem (1) is NP-hard due to its combinatorial nature [2]. Moreover, it is easy to notice that each feasible point of (1) is a local minimizer, which makes it difficult to find the global solution.

A popular way to solve the original problem (1) is to approximate it by replacing the  $\ell_0$  norm with the convex  $\ell_1$  norm

$$(P_1) : \begin{aligned} & \min_x \|x\|_1 \\ & \text{s.t. } y_0 = Ax. \end{aligned} \quad (2)$$

This work was supported in part by the  
\*\*\*\*\*, Grant \*\*\*\*\*  
\*\*\*\*\*, Grant \*\*\*\*\* and the  
\*\*\*\*\*.

Z.-L. Dong and Y.-H. Dai are with the State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China (e-mail: {zldong, dyh}@lsec.cc.ac.cn).

X.-Q. Yang is with the Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: mayangxq@polyu.edu.hk).

A variety of results have shown that if the RIC of matrix  $A$  satisfies certain condition, problem (2) can recover the optimal signal in (1) exactly [3]–[5].

In practice, it is more convinient to formulate the original problem (1) into a noise-aware version. Actually, we can only observe a noisy version  $y = y_0 + w$ , where  $w \in \mathbb{R}^n$  is the noise, and  $\|w\|_2 \leq \epsilon$ . Then we obtain a new problem

$$(P_{0,\sigma}) : \begin{aligned} & \min_x \|x\|_0 \\ & \text{s.t. } \|y - Ax\|_2 \leq \sigma, \end{aligned} \quad (3)$$

where  $\sigma$  is the estimation of the noise level  $\epsilon$ . A similar convexification strategy for (3) leads to

$$(P_{1,\sigma}) : \begin{aligned} & \min_x \|x\|_1 \\ & \text{s.t. } \|y - Ax\|_2 \leq \sigma. \end{aligned} \quad (4)$$

Researchers in [3], [4], [6] have shown that there is a stable recovery bound for the solution to (4), which is defined by the RIC and the noise level  $\epsilon$ , as well as the sparsity level  $N$ .

All the above results depend on RIC. However, as is well known, the computation of RIC for a given matrix is difficult [7]. In view of this, David L. Donoho and his collaborators proposed another way to show the stable recovery results for the sparse signal recovery problem constrained by a noise-aware overcomplete system [8]. They established the stable recovery bound defined by the *mutual coherence* constant of the dictionary, the sparsity level  $N$ , as well as the noise level  $\epsilon$ . Assume that the dictionary  $A = [A_1, A_2, \dots, A_n]$  has normalized columns under  $l^2$ -norm, which means  $\|A_i\|_2 = 1, \forall i = 1, 2, \dots, n$ ,  $A_i$  is the  $i$ -th column of matrix  $A$ . The *mutual coherence* constant is defined as

$$M = M(A) = \max_{1 \leq i, j \leq n, i \neq j} A_i^T A_j, \quad (5)$$

thus  $0 \leq M \leq 1$ . Smaller  $M$  means the dictionary is more incoherent. Compared with RIC, we can see that  $M$  is much easier to compute for a given matrix  $A$ .

Based on the above definition, the authors in [8] established the following results:

1. Suppose  $x_{0,\sigma}^*$  is the optimal solution to problem (3), if  $N < (M^{-1} + 1)/2$ , then

$$\|x_{0,\sigma}^* - x_0\|_2 \leq C_0(M, N) * (\epsilon + \sigma), \quad \forall \sigma \geq \epsilon > 0, \quad (6)$$

where the coefficient  $C_0(M, N) = \sqrt{1/(1 - M(2N - 1))}$ .

2. Suppose  $x_{1,\sigma}^*$  is the optimal solution to problem (4), if  $N < (M^{-1} + 1)/4$ , then

$$\|x_{1,\sigma}^* - x_0\|_2 \leq C_1(M, N) * (\epsilon + \sigma), \quad \forall \sigma \geq \epsilon > 0, \quad (7)$$

where the coefficient  $C_1(M, N) = \sqrt{1/(1 - M(4N - 1))}$ .

At the same time, researchers have also found that, in some cases, the non-convex model could be more efficient in sparse signal recovery problems compared to (4) [9]–[13]. There are a lot of works considering the sharp RIC condition for exact recovery of noiseless  $\ell_q$  ( $0 < q < 1$ ) minimization problem, as well as the stable recovery bound for noise-aware  $\ell_q$  ( $0 < q < 1$ ) minimization problem [14]–[16]. These stable recovery bound has the same difficulty in computing the RIC of a given dictionary. Thus, following Donoho's work, we want to give a similar stable recovery bound for  $\ell_q$  ( $0 < q < 1$ ) minimization problem without relying on RIC.

The main contribution of this letter is that we propose a unified stable recovery bound for the non-convex noise-aware model in compressed sensing. Specifically, consider using a non-convex relaxation strategy for  $P_{0,\sigma}$ , which replaces the  $\ell_0$  norm with the  $\ell_q$  ( $0 < q \leq 1$ ) quasi-norm. Minimize the  $\ell_q$  quasi-norm of the coefficient vector, subject to a noise-aware linear system

$$(P_{q,\sigma}) : \begin{aligned} & \min_x \|x\|_q \\ & \text{s.t. } \|y - Ax\|_2 \leq \sigma, \end{aligned} \quad (8)$$

where  $\|x\|_q = (\sum |x_i|^q)^{1/q}$ ,  $0 < q \leq 1$ . Denote the solution set of problem (8) as  $\mathcal{X}$ , we present a unified estimation about the upper bound of the error  $\|x_{q,\sigma}^* - x_0\|_2, \forall x_{q,\sigma}^* \in \mathcal{X}$ . If  $N < \frac{\gamma^{2/q-2}(M+1)}{4^{1/q}M}$ , then

$$\|x_{q,\sigma}^* - x_0\|_2 \leq C_q(M, N, \gamma)(\epsilon + \sigma), \forall x_{q,\sigma}^* \in \mathcal{X}, \quad (9)$$

where  $C_q(M, N, \gamma) = \sqrt{1/(1 - M(\frac{4^{1/q}}{\gamma^{2/q-2}}N - 1))}$ .  $M$  and  $N$  are the same as former definitions,  $\gamma$  is a constant lies in the interval  $[1/N, n/N]$ , which is defined later. Further analysis shows that, when  $\gamma > 2$ , the  $\ell_q$  ( $0 < q < 1$ ) model outperforms the  $\ell_1$  model in the sense that the former one has a smaller stable recovery bound, and the requirement of the sparsity  $N$  is weaker than the latter one.

The rest of this letter is organized as follows. Section II presents several basic lemmas that are needed in subsequent analysis. In Section III, we show the main theorem and some analysis about the result. Section IV concludes the letter and makes some comments on future work.

## II. SOME BASIC LEMMAS

First, we recall a well-known result without showing the proof [17].

*Lemma 2.1:*  $\|x\|_p \leq \|x\|_q, \forall x \in \mathbb{R}^n, 0 < p \leq q$ .

With this result, we can prove an important lemma in this letter.

*Lemma 2.2:*  $|a + b|^q + |b|^q \geq |a|^q, \forall a, b \in \mathbb{R}, 0 < q \leq 1$ .

*Proof:* It is obvious that

$$(|a + b| + |b|)^q \geq |a|^q, \quad (10)$$

thus, we only need to prove

$$|a + b|^q + |b|^q \geq (|a + b| + |b|)^q. \quad (11)$$

Due to lemma (2.1), we denote a vector  $x = (|a + b|, |b|)$  and  $p = 1$ , then the lemma holds. ■

Here is a lemma showing the relationship between the  $\ell_1$  norm and the  $\ell_q$  quasi-norm of error vector  $e$ .

*Lemma 2.3:* Let  $e = x_{q,\sigma}^* - x_0$ , then

$$(\gamma N)^{1/q-1} \|e\|_1 \leq \|e\|_q \leq n^{1/q-1} \|e\|_1, \forall 0 < q \leq 1, \quad (12)$$

where  $N = \|x_0\|_0$ ,  $\gamma$  lies in the interval  $[1/N, n/N]$ ,  $n$  is the dimension of  $e$ .

*Proof:* We first prove the second inequality. Based on the definition, we know that

$$\begin{aligned} \|e\|_q^q &= \sum_{i=1}^n |e_i|^q \\ &= \sum_{i=1}^n |e_i|^q \cdot 1 \\ &\leq \left( \sum_{i=1}^n (|e_i|^q)^{\frac{1}{q}} \right)^q \left( \sum_{i=1}^n 1^{\frac{1}{1-q}} \right)^{1-q} \\ &= \|e\|_1^q n^{1-q} \end{aligned} \quad (13)$$

where the inequality holds due to the Holder's inequality, thus  $\|e\|_q \leq n^{1/q-1} \|e\|_1$ .

From lemma (2.1), we know that  $\|e\|_1 \leq \|e\|_q, \forall 0 < q \leq 1$ , thus  $\gamma \geq 1/N$ . When  $e$  has only one nonzero entry, then the first equality in the lemma holds for any  $q \in (0, 1]$  if and only if  $\gamma = 1/N$ . From the second inequality, we know that  $\gamma \leq n/N$ . When  $e$  has all equal nonzero entries, then the first equality in the lemma holds for any  $q \in (0, 1]$  if and only if  $\gamma = n/N$ . Thus  $\gamma$  lies in the interval  $[1/N, n/N]$ . ■

Note that  $\gamma$  here is a measure of the sparsity level of the error vector  $e$ . In general,  $\gamma$  increases as the number of nonzero entries in  $e$  increases.

We end this section with the following well-known fact.

*Lemma 2.4:* For any vector  $x \in \mathbb{R}^n$ , we have

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2. \quad (14)$$

## III. MAIN RESULTS AND THE ANALYSIS

In this section, we present the main result of the stable recovery bound for the noise-aware non-convex minimization problem (8). This result extends Donoho and his coauthors' result in [8] to the general case where  $q \in (0, 1]$ . We first present the main theorem and the details of the proof and then make our comments.

### A. Main results

Before the main theorem, we show a lemma about the optimal solutions to problem (8).

*Lemma 3.1:* For the optimal solution to problem (8), the inequality in the constraint is active.

*Proof:* In order to prove this result, we rewrite the problem (8) into an equivalent form

$$\begin{aligned} & \min_x \|x\|_q^q \\ & \text{s.t. } \|y - Ax\|_2^2 \leq \sigma^2, \\ & 0 < q \leq 1. \end{aligned} \quad (15)$$

The optimal solution to problem (15) also solves problem (8). Assume that  $x_{q,\sigma}^*$  is a local minimizer of problem (15), and the support set of  $x_{q,\sigma}^*$  is  $\mathcal{S}$ ,  $x_{q,\sigma}^*$  is also the stationary point. Then the first-order optimality conditions for this problem holds if there exists a pair of  $(x_{q,\sigma}^*, \lambda^*)$  such that

$$\begin{aligned} \nabla(\|(x_{q,\sigma}^*)_{\mathcal{S}}\|_q^q) - \lambda^* A_{\mathcal{S}}^T(y_{\mathcal{S}} - A_{\mathcal{S}}(x_{q,\sigma}^*)_{\mathcal{S}}) &= 0, \\ \lambda^*(\|y - Ax_{q,\sigma}^*\|_2^2 - \sigma^2) &= 0, \\ \|y - Ax_{q,\sigma}^*\|_2^2 &\leq \sigma^2, \\ \lambda^* &\geq 0. \end{aligned} \quad (16)$$

From the first equality in (16), we know that if  $\lambda^* = 0$ , then  $x_{q,\sigma}^*$  should be a zero vector. The zero vector may not satisfy the third inequality in the KKT system (16), since  $y$  is a nonzero observation vector with noise. Thus, we do accept the fact that  $\lambda^* > 0$ . And then according to the complementarity condition, we will have  $\|y - Ax_{q,\sigma}^*\|_2^2 = \sigma^2$ , i.e., the conclusion of the lemma holds. ■

We are now ready to present our main result.

**Theorem 3.2:** Suppose the ideal sparse representation signal satisfies

$$N := \|x_0\|_0 < \frac{\gamma^{2/q-2}(M+1)}{4^{1/q}M}, \quad (17)$$

where  $M$  is defined in (5) and  $\gamma$  is defined in Lemma (2.3). Then the difference between any solution to  $P_{q,\sigma}$  and  $x_0$ , assuming  $\sigma \geq \epsilon$ , is upper bounded

$$\|x_{q,\sigma}^* - x_0\|_2^2 \leq \frac{(\epsilon + \sigma)^2}{1 - M(\frac{4^{1/q}}{\gamma^{2/q-2}}N - 1)}, \forall x_{q,\sigma}^* \in \mathcal{X}. \quad (18)$$

*Proof:* Let  $x^*$  be the optimal solution to problem (8), denote the error as  $e = x^* - x_0$ , the process to estimate the bound can be formulated into an optimization problem in this form

$$\begin{aligned} \max \quad & \|e\|_2^2 \\ \text{s.t.} \quad & \mathcal{X} := \arg \min_x \{\|x\|_q^q : \|Ax - y\|_2 \leq \sigma\}, \\ & x^* \in \mathcal{X}, \\ & y = Ax_0 + w, \\ & \|w\|_2 \leq \epsilon, \|x_0\|_0 = N. \end{aligned} \quad (19)$$

The constraints indicate that we consider the worst case since all the optimal solutions to problem (8) are taken into consideration. Define the optimal function value of (19) as  $\text{Val}(19)$ . The upper bound of  $\text{Val}(19)$  is the target, and the main idea of the proof is to expand the feasible set of problem (19) sequentially and obtain a series of upper bound on  $\text{Val}(19)$ .

Note that  $x_0$  itself is a feasible point of the constraint  $\|Ax - y\|_2 \leq \sigma$ . Based on the conclusion of Lemma (3.1), do some simple substitution and eliminate  $y$ , we can relax the constraints in (19) to

$$\left\{ e \mid \begin{aligned} & \|x_0 + e\|_q^q \leq \|x_0\|_q^q, \|Ae - w\|_2 = \sigma \\ & \|w\|_2 \leq \epsilon, \|x_0\|_0 = N \end{aligned} \right\}. \quad (20)$$

Define  $\mathcal{S} = \{i | (x_0)_i \neq 0\}$  as the support set and the complement set as  $\mathcal{S}^c = \{1, 2, \dots, n\} \setminus \mathcal{S}$ . From Lemma (2.2)

we know that

$$\|x_0 + e\|_q^q - \|x_0\|_q^q \geq \|e_{\mathcal{S}^c}\|_q^q - \|e_{\mathcal{S}}\|_q^q = \|e\|_q^q - 2 \sum_{i \in \mathcal{S}} |e_i|^q. \quad (21)$$

Therefore, we can relax the feasible set further and rewrite the problem (19) as

$$\begin{aligned} \max_{e, \mathcal{S}, w} \quad & \|e\|_2^2 \\ \text{s.t.} \quad & \|e\|_q^q \leq 2 \sum_{i \in \mathcal{S}} |e_i|^q, \\ & \|Ae - w\|_2 = \sigma, \\ & \|w\|_2 \leq \epsilon, \#\mathcal{S} = N, \end{aligned} \quad (22)$$

which has a larger optimal function value than (19).

Using the fact that  $\|Ae - w\|_2 \geq \|Ae\|_2 - \|w\|_2$ , we can relax the feasible set in (22) and obtain

$$\begin{aligned} \max_{e, \mathcal{S}} \quad & \|e\|_2^2 \\ \text{s.t.} \quad & \|e\|_q^q \leq 2 \sum_{i \in \mathcal{S}} |e_i|^q, \\ & \|Ae\|_2 \leq \Sigma, \#\mathcal{S} = N, \end{aligned} \quad (23)$$

where  $\Sigma = \epsilon + \sigma$ .

Now it's turn to deal with the constraint  $\|Ae\|_2 \leq \Sigma$ . Recall the definition of the *mutual coherence*,  $M = M(A) = \max_{i \neq j} |G(i, j)|$ , where  $G$  is the Gram matrix  $G = A^T A$ . For any matrix  $X$ , denote  $|X|$  as the matrix after taking absolute values for all the entries in  $X$ , this include the vector and the constant as the special case. Denote  $\mathbf{1}$  as a square matrix with all entries equal to one, and the dimension is correspondingly adaptive. Then we can relax the constraint as follows

$$\begin{aligned} \Sigma^2 &\geq \|Ae\|_2^2 = e^T G e = \|e\|_2^2 + e^T (G - \mathbf{I}) e \\ &\geq \|e\|_2^2 - |e^T| G - \mathbf{I} |e| \\ &\geq \|e\|_2^2 - M |e^T| \mathbf{1} - \mathbf{I} |e| \\ &= (1 + M) \|e\|_2^2 - M \|e\|_1^2 \\ &\geq (1 + M) \|e\|_2^2 - \frac{M}{(\gamma N)^{2/q-2}} \|e\|_q^2, \end{aligned} \quad (24)$$

where the last inequality is due to Lemma (2.3). Thus, we can obtain the following problem with a larger optimal function value

$$\begin{aligned} \max_{e, \mathcal{S}} \quad & \|e\|_2^2 \\ \text{s.t.} \quad & \|e\|_q^q \leq 2 \sum_{i \in \mathcal{S}} |e_i|^q, \\ & (1 + M) \|e\|_2^2 - \frac{M}{(\gamma N)^{2/q-2}} \|e\|_q^2 \leq \Sigma^2, \#\mathcal{S} = N. \end{aligned} \quad (25)$$

Now we can see that the target term of the objective function appears in the second constraint. Our task is to replace the other terms in the constraints with the target term and obtain an upper bound for the function value. First, we separate the error vector into two parts,  $e = (e_{\mathcal{S}}, e_{\mathcal{S}^c})$ , where  $e_{\mathcal{S}} = \{e_i | i \in \mathcal{S}\}$  and  $e_{\mathcal{S}^c} = \{e_i | i \in \mathcal{S}^c\}$ . Then it's obvious that

$$\begin{aligned} \|e\|_2^2 &= \|e_{\mathcal{S}}\|_2^2 + \|e_{\mathcal{S}^c}\|_2^2, \\ \|e\|_q^2 &= (\|e_{\mathcal{S}}\|_q^q + \|e_{\mathcal{S}^c}\|_q^q)^{2/q}. \end{aligned} \quad (26)$$

Based on Lemma(2.3) and (2.4), we have

$$\begin{aligned} \|e_S\|_2 &\leq \|e_S\|_q \leq N^{(1/q-1/2)} \|e_S\|_2, \\ \|e_{S^c}\|_2 &\leq \|e_{S^c}\|_q \leq (n-N)^{(1/q-1/2)} \|e_{S^c}\|_2. \end{aligned} \quad (27)$$

In order to make the symbol simpler, we define the following intermediate variables

$$\begin{aligned} V_S &= \|e_S\|_q^q, V_{S^c} = \|e_{S^c}\|_q^q, \\ \eta_S &= \|e_S\|_2^2 / \|e_S\|_q^2, \eta_{S^c} = \|e_{S^c}\|_2^2 / \|e_{S^c}\|_q^2, \end{aligned} \quad (28)$$

and reformulate the optimization problem (25) into an optimization problem on  $(V_S, V_{S^c}, \eta_S, \eta_{S^c}) \in R^4$

$$\begin{aligned} \max \quad & \eta_S V_S^{2/q} + \eta_{S^c} V_{S^c}^{2/q} \\ \text{s.t.} \quad & (1+M)(\eta_S V_S^{2/q} + \eta_{S^c} V_{S^c}^{2/q}) \\ & - \frac{M}{(\gamma N)^{2/q-2}} (V_S + V_{S^c})^{2/q} \leq \Sigma^2, \\ & 0 \leq V_{S^c} \leq V_S, N^{1-2/q} \leq \eta_S \leq 1, 0 < \eta_{S^c} \leq 1. \end{aligned} \quad (29)$$

Define the ratio between  $V_S$  and  $V_{S^c}$  as  $\tau = V_{S^c}/V_S$ , then we have  $0 \leq \tau \leq 1$ . (29) can be reformulated as

$$\begin{aligned} \max \quad & (\eta_S + \eta_{S^c} \tau^{2/q}) V_S^{2/q} \\ \text{s.t.} \quad & (1+M)(\eta_S + \eta_{S^c} \tau^{2/q}) V_S^{2/q} \\ & - \frac{M}{(\gamma N)^{2/q-2}} (1+\tau)^{2/q} V_S^{2/q} \leq \Sigma^2, \\ & V_S \geq 0, N^{1-2/q} \leq \eta_S \leq 1, 0 < \eta_{S^c} \leq 1, 0 \leq \tau \leq 1. \end{aligned} \quad (30)$$

Let  $\mu = (1+\tau)^{2/q} / (\eta_S + \eta_{S^c} \tau^{2/q})$ , then it is easy to see that  $1 \leq \mu \leq 4^{1/q} N^{2/q-1}$  under the constraints in (30). Denote the objective function value as  $V = V_S^{2/q} (\eta_S + \eta_{S^c} \tau^{2/q})$ , then the first constraint in (30) is equivalent to

$$(1+M)V - \frac{M}{(\gamma N)^{2/q-2}} \mu V \leq \Sigma^2. \quad (31)$$

Substitute the upper bound of  $\mu$  into this inequality, and to maintain the coefficient of  $V$  to be positive, we will have the requirement of sparsity  $N$  as follows

$$(1+M) - \frac{M}{(\gamma N)^{2/q-2}} \mu \geq (1+M) - M \frac{4^{1/q} N}{\gamma^{2/q-2}} > 0. \quad (32)$$

Thus, we have

$$N < \frac{\gamma^{2/q-2} (M+1)}{4^{1/q} M}, \quad (33)$$

and

$$V \leq \frac{\Sigma^2}{1 - M(\frac{\mu}{(\gamma N)^{2/q-2}} - 1)} \leq \frac{\Sigma^2}{1 - M(\frac{4^{1/q} N}{\gamma^{2/q-2}} - 1)}, \quad (34)$$

which concludes the proof. ■

#### B. On the parameter $\gamma$ in the recovery bound (18)

From the above results, we can see that, the parameter  $\gamma$  plays an important role in the recovery bound (18). We will remark that for which  $\gamma$  the recovery bound of model (8) is better than that of model (4).

**Case 1:** if  $\gamma = 2$ , we have  $C_q(M, N, \gamma) = C_1(M, N)$ . Thus the sparsity requirement in (8) is equal to that in (4), and the

recovery bound of model (8) is the same to that of model (4). This leads to the conclusion that the non-convex minimization model (8) approximates the original model (1) no worse than the convex minimization model (4).

**Case 2:** if  $\gamma > 2$ , we have  $C_q(M, N, \gamma) < C_1(M, N)$ . Thus the sparsity requirement in (8) is weaker than that in (4), and the recovery bound of model (8) is smaller than that of model (4). This means that the non-convex minimization model (8) is better than the convex minimization model (4) in approximating the original model (1).

For example, let  $q = 1/2$ . From Lemma (2.3), we have

$$\gamma N \|e\|_1 \leq \|e\|_{1/2}. \quad (35)$$

Since  $\gamma > 2$ , from simple calculation, we can see that (35) holds if

$$\max_{i=1,2,\dots,n} |e_i|^{1/2} \leq \frac{\sum_{j \neq i} |e_j|^{1/2}}{\gamma N - 1}, \quad (36)$$

i.e.

$$2 < \gamma \leq \frac{\sum_j |e_j|^{1/2}}{N * \max_{i=1,2,\dots,n} |e_i|^{1/2}}. \quad (37)$$

Condition (37) requires that the number of nonzero entries in error vector  $e$  should be larger than  $\gamma N$  in this case. This requirement seems to be very strong, since it says that the worst solution in the optimal solution set  $\mathcal{X}$  has at least  $N$  tails except the right sparse locations. However, recent numerical experiments indicate that this could be true in the noise-aware case; see for example [18]. Thus, it is conceivable that for the noise-aware models (4) and (8), (8) outperforms (4) in approximating (1) in the sense that (8) has a smaller worst-case stable recovery bound and weaker requirement for the sparsity of the ideal signal.

## IV. CONCLUSION

In this letter, we present a unified worst-case stable recovery bound for the noise-aware  $\ell_q$  ( $0 < q \leq 1$ ) minimization problem. The result reduces to the original result in Donoho's paper [8] when  $q = 1$ . We introduce a parameter  $\gamma$  which describes the relationship between the  $\ell_1$  norm and the  $\ell_q$  quasi-norm of the error vector  $e$ . Further analysis indicates that, for  $\gamma > 2$ , the non-convex  $\ell_q$  ( $0 < q < 1$ ) minimization model outperforms the convex  $\ell_1$  model, in the sense that, the former model has a smaller worst-case recovery bound and weaker sparsity requirement.

Moreover, the recovery bound for (8) proposed in this letter is very loose due to the unduely relaxation, similar to the recovery bound for (4) in [8]. Even so, the result in this letter still has some guiding significance. Since it is easier to compute the constant in the recovery bound (18) for a given problem. We are also looking forward to the extension of this result to the general sparse group sparse optimization problem [19], [20], which is also popular in the recent period.

## REFERENCES

- [1] Y.C. Eldar and G. Kutyniok. *Compressed sensing: theory and applications*. Cambridge University Press, 2012.
- [2] B.K. Natarajan. Sparse Approximate Solutions to Linear Systems. *SIAM J. COMPUT.*, 24(2):227–234, 1995.

- [3] T.T. Cai and A. Zhang. Sharp RIP Bound for Sparse Signal and Low-rank Matrix Recovery. *Appl. Comput. Harmon. A.*, 35(1):74–93, 2013.
- [4] E.J. Candès and T. Tao. Decoding by Linear Programming. *IEEE Trans. Inf. Theory*, 51(12):4203–4215, 2005.
- [5] D.L. Donoho. For Most Large Underdetermined Systems of Linear Equations the Minimal  $\ell_1$ -norm Solution is Also the Sparsest Solution. *Commun. Pure Appl. Math.*, 59(6):797–829, 2006.
- [6] E.J. Candès, J.K. Romberg, and T. Tao. Stable Signal Recovery from Incomplete and Inaccurate Measurements. *Commun. Pure Appl. Math.*, 59(8):1207–1223, 2006.
- [7] A.S. Bandeira, E. Dobriban, D.G. Mixon, and W.F. Sawin. Certifying the Restricted Isometry Property is Hard. *IEEE Trans. Inf. Theory*, 59(6):3448–3450, 2013.
- [8] D.L. Donoho, M. Elad, and V.N. Temlyakov. Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise. *IEEE Trans. Inf. Theory*, 52(1):6–18, 2006.
- [9] R. Chartrand. Restricted Isometry Properties and Nonconvex Compressive Sensing. *Inverse Problems*, 24(3):1–14, 2008.
- [10] S. Foucart and M.J. Lai. Sparsest Solutions of Underdetermined Linear Systems via  $\ell_q$ -minimization for  $0 < q \leq 1$ . *Appl. Comput. Harmon. A.*, 26(3):395–407, 2009.
- [11] R. Saab and Ö. Yilmaz. Sparse Recovery by Non-convex Optimization: Instance Optimality. *Appl. Comput. Harmon. A.*, 29(1):30–48, 2010.
- [12] N. Mourad and J.P. Reilly. Minimizing Nonconvex Functions for Sparse Vector Reconstruction. *IEEE Trans. on Signal Processing*, 58(7):3485–3496, 2010.
- [13] Y.F. Liu, S.Q. Ma Y.H. Dai, and S.Z. Zhang. A Smoothing SQP Framework for a Class of Composite  $L_q$  Minimization over Polyhedron. *Mathematical Programming*, 158(1):467–500, 2016.
- [14] R. Saab, R. Chartrand, and Ö. Yilmaz. Stable Sparse Approximations via Nonconvex Optimization. *ICASSP 2008, IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3885–3888, 2008.
- [15] J.M. Wen, D.F. Li, and F.M. Zhu. Stable Recovery of Sparse Signals via  $\ell_p$  Minimization. *Preprint, arXiv:1406.4328 [cs.IT]*, 2014.
- [16] C.B. Song and S.T. Xia. Sparse signal recovery by minimization under restricted isometry property. *IEEE Signal Processing Letters*, 21(9):1154–1158, 2014.
- [17] Y.H. Hu, C. Li, K.W. Meng, J. Qin, and X.Q. Yang. Group Sparse Optimization via  $\ell_{p,q}$  Regularization. *Preprint, arXiv:1601.07779v1 [math.OC]*, 2016.
- [18] X.J. Chen, Z.S. Lu, and T.K. Pong. Penalty Methods for a Class of Non-Lipschitz Optimization Problems. *To appear in SIAM J. Optim.*, 2016.
- [19] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani. A sparse-group lasso. *J. Comput. Graph. Stat.*, 22(2):231–245, 2013.
- [20] M. Vincent and N.R. Hansen. Sparse Group Lasso and High Dimensional Multinomial Classification. *Computational Statistics & Data Analysis*, 71:771–786, 2014.